

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas – ICEX
Departamento de Estatística
Especialização em Estatística

Rafaela da Silveira Pinto

**Efeito do plano amostral complexo em inquérito
epidemiológico de saúde bucal em Minas Gerais 2012**

Belo Horizonte

2019

Rafaela da Silveira Pinto

**Efeito do plano amostral complexo em inquérito
epidemiológico de saúde bucal em Minas Gerais 2012**

Versão final

Monografia de especialização apresentada ao Departamento de Estatística da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Especialista em Estatística.

Orientador: Professor Thiago Rezende dos Santos

Co-orientador: Professor Mauro Henrique Nogueira Guimarães de Abreu

Belo Horizonte

2019

© 2019, Rafaela da Silveira Pinto.

Todos os direitos reservados

Ficha catalográfica elaborada pela Biblioteca do ICEX - UFMG

Pinto, Rafaela da Silveira.

P659e Efeito do plano amostral complexo em inquérito epidemiológico de saúde bucal em Minas Gerais 2012 / Rafaela da Silveira Pinto. — Belo Horizonte, 2019.

53 f.: il.; 29 cm.

Monografia (especialização) - Universidade Federal de Minas Gerais – Departamento de Estatística.

Orientador Thiago Rezende dos Santos.

Coorientador: Mauro Henrique Nogueira Guimarães de Abreu.

1. Estatística. 2. Saúde bucal – Minas Gerais.
3. Epidemiologia – Métodos estatísticos. 4. Análise por Conglomerados. I. Orientador. II. Coorientador.
III. Título.

CDU 519.2(043)

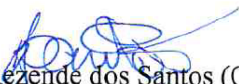


Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
 Av. Pres. Antônio Carlos, 6627 - Pampulha
 31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br
 Tel: 3409-5923 – FAX: 3409-5924

ATA DO 199º. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DE RAFAELA DE SILVEIRA PINTO.

Aos nove dias do mês de outubro de 2019, às 14:00 horas, na Sala 2076 do Instituto de Ciências Exatas, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística, para julgar a apresentação do trabalho de fim de curso da aluna **Rafaela da Silveira Pinto**, intitulado: **“Efeito do Plano Amostral (EPA) em inquéritos epidemiológicos de saúde bucal em Minas Gerais 2012”**, como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, o Presidente da Comissão, Professor Thiago Rezende dos Santos – Orientador, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra à candidata para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa da candidata. Após a defesa, os membros da banca examinadora reuniram-se sem a presença da candidata e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: a candidata foi considerada Aprovada condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje por unanimidade. O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 09 de outubro de 2019.


 Prof. Thiago Rezende dos Santos (Orientador)
 Departamento de Estatística / UFMG


 Prof. Mauro Henrique Nogueira Guimarães Abreu (Co orientador)
 Faculdade de Odontologia / UFMG


 Prof.ª Lourdes Coral Montenegro
 Departamento de Estatística / UFMG


 Prof.ª Andrea Clemente Palmier
 Faculdade de Odontologia / UFMG

Agradecimentos

Agradeço primeiramente à Deus, por sempre prover a força necessária. Agradeço aos professores do curso e à secretaria administrativa por se disporem a dedicar seus sábados na formação dos alunos. Agradeço aos colegas de curso, em especial à Cintia, Fabrícia, Mirele e Marcela, companheiras de sábados e almoços no “bandejão”. Agradeço aos meus familiares, que sempre me apoiam, mas que, a cada curso que digo que vou fazer, perguntam-me se eu não vou parar de estudar... acho que a resposta é NÃO! Agradeço imensamente ao meu marido Alex, que está sempre ao meu lado, dividindo os momentos felizes e apoiando nos momentos de sufoco! Por último e não menos importante, agradeço aos meus orientadores, que me deram suporte e motivação quando eu estava 99% certa que iria desistir.

A todos vocês, muito obrigada!

Resumo

OBJETIVOS: O objetivo deste trabalho foi apresentar os aspectos mais importantes do plano amostral da Pesquisa das condições de saúde bucal da população mineira - SB Minas Gerais, realizado em 2012, especialmente no que se refere ao Efeito do Plano Amostral (EPA).

MÉTODOS: As estimativas foram calculadas para cada domínio definido segundo unidade geográfica e grupos demográficos, ou seja, para as idades de 5 e 12 anos e as faixas etárias de 15 a 19, 35 a 44, 65 a 74 anos e nos diferentes domínios da pesquisa SB Minas Gerais 2012 e para Capital, Interior I e Interior II. A comparação dessas estimativas com ou sem pesos básicos permite avaliar o efeito da homogeneidade intraclasse e do impacto dos pesos amostrais na precisão.

RESULTADOS: De modo geral, nota-se baixo impacto do processo de conglomerados e do peso amostral na precisão quando comparadas as estimativas com e sem a ponderação nas prevalências de cárie estimadas. As estimativas para médias, erros padrão e EPA do CPOD nos grupos de 15 a 19 anos e de 35 a 44 anos também podem ser consideradas estáveis para a grande maioria dos domínios e, assim, compatíveis com os critérios de precisão previstos no delineamento. Entretanto, o maior valor para o EPA foi obtido no grupo de 35 a 44 anos para municípios do Interior II, ainda assim não afetando fortemente a estimativa pontual, apenas aumentando a amplitude do intervalo de confiança e o erro padrão. Os EPA apresentados consolidam os resultados obtidos para médias e proporções nos três domínios geográficos e em 4 domínios etários (excetuando-se a faixa etária de adultos para o Interior II).

CONCLUSÕES: As análises estatísticas realizadas com o banco de dados proveniente de amostras complexas devem levar em consideração o delineamento amostral. Além disso, em estudos epidemiológicos futuros, ao se planejar a amostragem, deve ser levada em consideração a necessidade de variar o EPA de acordo com as faixas etárias.

Palavras chave: Inquéritos de Saúde Bucal, métodos. Amostragem por Conglomerados. Projetos de Pesquisa Epidemiológica.

Abstract

OBJECTIVES: The objective of this paper was to present the most important aspects of the sampling plan of the Survey of oral health conditions of the Minas Gerais' population - SB Minas Gerais, conducted in 2012, especially regarding the design effect (deff). **METHODS:** Estimates were calculated for each domain defined by geographical units (Capital, Interior I and Interior II) and demographic groups by 5 and 12 years and age groups 15 to 19, 35 to 44, 65 to 74 years. Comparing these estimates with or without basic weights allows us to evaluate the effect of intraclass homogeneity and the impact of sample weights on precision. **RESULTS:** Overall, there is a low impact of clustering and sample weight on precision when comparing estimates with and without weighting on estimated caries prevalence. Estimates for mean, standard errors, and deff of the DMFT index in the 15-19 age group and 35-44 age group can also be considered stable for the vast majority of domains and thus compatible with the precision criteria predicted in the design. However, the highest value for the deff was obtained in the group of 35 to 44 years for municipalities of the Interior II, yet not strongly affecting the point estimate, only increasing the confidence interval amplitude and standard error. The presented deff consolidate the results obtained for averages and proportions in the three geographic domains and in four age domains (except for the adults in the Interior II domain). **CONCLUSIONS:** Statistical analyzes performed with the database from complex samples should take into account the sample design. Furthermore, in future epidemiological studies, when planning sampling, consideration should be given to the need to vary the deff according to age groups.

Keywords: Dental health surveys. Sampling studies. Cross sectional studies.

Lista de Ilustrações

Figura 1 - Distribuição dos municípios mineiros segundo quartis (fator de alocação).	29
Figura 2 - Distribuição dos municípios mineiros para composição dos domínios da pesquisa.	30
Figura 3 - Municípios sorteados de acordo com o domínio.	30
Figura 4 - Fluxograma para ajuste do banco de dados dos setores censitários.	32
Quadro 1 - Características dos inquéritos nacionais de saúde bucal.	13
Quadro 2 - Classificação do Fator de Alocação segundo quartis.	29
Quadro 3 - Condições e idades/grupos etários a serem pesquisados.	33

Lista de tabelas

Tabela 1 - Estimativas de proporção de indivíduos com cárie, erros padrão e efeito de desenho com e sem ponderação para os grupos etários de acordo com os domínios de estudo. SB Minas Gerais, 2012..... 20

Tabela 2 - Estimativas para o ceo-d/CPOD (dentes cariados, perdidos e obturados), erros padrão e efeito de desenho com e sem ponderação para os grupos etários de acordo com os domínios de estudo. SB Minas Gerais, 2012..... 20

Tabela 3 - Estimativas de proporção de indivíduos com cárie, erros padrão e efeito de desenho com e sem ponderação de acordo com os domínios de estudo. SB Minas Gerais, 2012..... 21

Tabela 4 - Estimativas para o ceo-d/CPOD (dentes cariados, perdidos e obturados), erros padrão e efeito de desenho com e sem ponderação de acordo com os domínios de estudo. SB Minas Gerais, 2012..... 21

Lista de abreviaturas e siglas

AAS – Amostragem aleatória simples

AASc – Amostragem aleatória simples com reposição

ceo-d – Número de dentes cariados, perdidos e obturados na dentição decídua

CPOD – Número de dentes cariados, perdidos e obturados na dentição permanente

Deff – Design effect

IC95% - Intervalo de confiança a 95%

EPA – Efeito do plano amostral

IBGE – Instituto Brasileiro de Geografia e Estatística

MS – Ministério da Saúde

PNAD – Pesquisa nacional por amostragem de domicílios

PME – Pesquisa mensal de emprego

PAC – Plano amostral complexo

PPT – Probabilidade proporcional ao tamanho

SB Brasil 2003 - Pesquisa das condições de saúde bucal da população brasileira

SB Brasil 2010 – Pesquisa nacional de saúde bucal

SB Minas Gerais 2012 – Pesquisa das condições de saúde bucal da população mineira

SESI – Serviço Social da Indústria

UPA – Unidade primária de amostragem

Sumário

1	INTRODUÇÃO.....	11
1.1	Sobre planos amostrais complexos.....	11
1.2	O Projeto SB Minas Gerais.....	12
2	REVISÃO DE LITERATURA.....	14
3	OBJETIVO.....	16
4	METODOLOGIA.....	17
5	RESULTADOS.....	19
6	DISCUSSÃO.....	22
7	CONSIDERAÇÕES FINAIS.....	24
	REFERÊNCIAS.....	25
	ANEXO 1 – Detalhamento do plano amostral da do Projeto SB Minas Gerais – Pesquisa das Condições de Saúde Bucal da População Mineira.....	28
1	Área e População de Estudo - SB Minas Gerais 2012.....	28
2	Características do Projeto SB Minas Gerais - Pesquisa das condições de saúde bucal da população mineira.....	28
3	Plano Amostral.....	31
3.1	Domínios e Unidades Primárias de Amostragem (UPA).....	31
3.2	Unidades Secundárias de Amostragem (USA): Setores Censitários.....	31
3.3	Idades-índices e grupos etários do Projeto SB Minas Gerais.....	32
	ANEXO 2 - Número de domicílios e indivíduos pesquisados e respectivas taxas de resposta de acordo com domínio e grupo etário.....	35
	APÊNDICE – Programação e resultados das análises no R.....	36
1	Capital.....	36
1.1	Com plano amostral complexo e com pesos.....	36
1.2	Com plano amostral complexo e sem pesos.....	38
1.3	Sem peso, sem estrato e cluster (Amostra Aleatória Simples com reposição).....	40
2	Interior I.....	42
2.1	Com plano amostral complexo e com pesos.....	42
2.2	Com plano amostral complexo e sem pesos.....	44
2.3	Sem peso, sem estrato e cluster (Amostra Aleatória Simples com reposição).....	46
3	Interior II.....	48
3.1	Com plano amostral complexo e com pesos.....	48
3.2	Com plano amostral complexo e sem pesos.....	50
3.3	Sem peso, sem estrato e cluster (Amostra Aleatória Simples com reposição).....	52

1 INTRODUÇÃO

1.1 Sobre planos amostrais complexos

Planos amostrais complexos vem sendo utilizados em diversos estudos, de diferentes áreas (IBGE, 2016; SILVA *et al.*, 2002; LISA *et al.*, 2013; GLENN *et al.*, 2019; BELL *et al.*, 2012), quando não se dispõe de uma lista de indivíduos para sorteio de uma amostra. Métodos para estimar o efeito do desenho em dados de dois níveis são bem estabelecidos, porém, a complexidade analítica dos dados de três níveis surge da necessidade de entender as relações entre e dentro dos níveis de estratificação da amostra e da especificação apropriada da estrutura de variância. Nos últimos anos, com o aumento da oferta de recursos computacionais houve um aumento da literatura disponível sobre questões técnicas e estatísticas relacionadas a dados de três níveis, todavia, as questões mais problemáticas continuam sendo a estimativa do poder estatístico, do erro amostral e do tamanho da amostra. Alguns desafios relacionados a este fato são: Primeiro, o efeito de aninhamento multinível introduz mais de uma correlação intracluster no modelo. Em segundo, a variância apresentada pelos desfechos são mais complexas que o de um design de nível único ou de dois níveis. Em terceiro lugar, três cálculos amostrais são necessários - um para cada nível - e cada tamanho de amostra afeta o erro padrão das estimativas de forma diferente (CUNNINGHAM; JOHNSON, 2012).

No caso da Pesquisa das condições de saúde bucal da população mineira (SB Minas Gerais 2012), assim como na Pesquisa Nacional de Saúde Bucal (SB Brasil 2010), o modelo geral usado foi o da amostragem por conglomerados em múltiplos estágios, nos quais as unidades amostrais foram sorteadas com probabilidade proporcional ao tamanho (PPT). No primeiro estágio, sortearam-se 30 municípios para o interior de cada região (Interior I e II). São essas as unidades primárias de amostragem (UPA) que foram incluídas na fase de elaboração dos arquivos, bem como nos cálculos das estimativas de erros padrão e intervalos de confiança. No segundo estágio, foram sorteados quatro setores censitários nos municípios que compunham as amostras do interior. Cada região geográfica contou com 120 setores. No terceiro estágio, apanharam-se aleatoriamente domicílios em cada setor sorteado no estágio anterior por meio da folha de arrolamento (PINTO *et al.*, 2018). Maior detalhamento do plano amostral está disponível no tópico a seguir.

Como as unidades amostrais nos níveis mais baixos de um estudo podem não ser independentes entre si, o pressuposto de tamanho de amostra aleatória simples não é válido. Por exemplo, os

pacientes que visitam o mesmo médico podem compartilhar características comuns que variam de pacientes que visitam um médico diferente. Além disso, as características dos médicos dentro do mesmo centro de saúde tendem a ser homogênea, enquanto os grupos médicos entre os centros são heterogêneos (CUNNINGHAM; JOHNSON, 2012). A literatura existente sugere que uma maneira de explicar o efeito dessa dependência intracluster é inflar as estimativas do erro padrão por um efeito de agrupamento, também conhecido como efeito do plano amostral ou efeito de design (em inglês, design effect – *deff*) (CUNNINGHAM; JOHNSON, 2012).

Usando o SB Minas Gerais como exemplo, analogamente ao exemplo apresentado no parágrafo anterior, moradores de uma mesma cidade tem contextos próximos do que aqueles que não moram na mesma cidade. Da mesma forma, ocorre com setores censitários e domicílios.

1.2 O Projeto SB Minas Gerais

A relação entre epidemiologia e saúde pública tem se estabelecido de forma muito próxima. Historicamente, a epidemiologia é considerada a principal ferramenta para subsidiar os profissionais e administradores de saúde. Muitas outras áreas de estudos também são importantes para a saúde pública, no entanto, a epidemiologia exerce um papel fundamental por causa do seu foco populacional e métodos (GOUDA; POWLES, 2014).

O SB Minas Gerais 2012 (MINAS GERAIS, 2013) seguiu a metodologia do SB Brasil 2010 (BRASIL, 2010) tanto no que se refere às idades analisadas e aos índices epidemiológicos aferidos quanto no modelo amostral complexo e, da mesma forma que esta última, é uma das estratégias de vigilância em saúde que utilizam dados primários para gerar informações relevantes que subsidiem a implementação de políticas públicas de saúde bucal. O SB Minas Gerais 2012 foi o primeiro inquérito de saúde bucal de base estadual com amostra representativa. No ano de 2003, concomitantemente com o SB Brasil 2003, a gestão estadual incluiu outros municípios na amostra original do inquérito nacional realizado nesta ocasião (amostra de conveniência), não sendo formatado para ser representativo.

O SB Minas Gerais 2012 foi idealizado ainda durante a condução do SB Brasil 2010 uma vez que este não dispunha de amostras representativas para os estados e, então, durante o ano de 2011, foram realizadas as reuniões de planejamento para que a coleta ocorresse no ano de 2012 (MINAS GERAIS, 2012).

No Brasil já foram realizados cinco grandes inquéritos epidemiológicos de saúde bucal os quais detalhamentos as principais características no quadro a seguir:

Quadro 1 - Características dos inquéritos nacionais de saúde bucal.

Ano	Instituição	Amostra	Problemas investigados	Faixa Etária (Anos)
1986	Ministério da Saúde-MS	n=25.407 16 capitais Escolas públicas e privadas Zona Urbana	Cárie Doença periodontal Uso e necessidade de prótese total	6-12 15-19 25-44 50-59
1993	Serviço Social da Indústria-SESI	n=110.640 21 estados e DF (capitais e interior) 114 municípios Escolas públicas e do SESI Zona Urbana	Cárie	3-14
1996	Ministério da Saúde-MS	n=30.240 27 capitais e DF Escolas públicas e privadas Zona Urbana	Cárie	6-12
2002-2003	Ministério da Saúde-MS	n=108.921 Capitais e Interior (250 municípios) Escolas públicas e privadas e domicílios Zona Urbana e Rural	Cárie Doença periodontal Fluorose Má-oclusão Uso e necessidade de prótese Alteração de tecido mole Acesso a serviços odontológicos Autopercepção em saúde bucal Condições socioeconômicas	18-36 meses 5 12 35-44 65-74
2010	Ministério da Saúde-MS	n=36.904 Capitais e Interior (177 municípios)	Cárie Doença periodontal Fluorose Má-oclusão Uso e necessidade de prótese Alteração de tecido mole Acesso a serviços odontológicos Autopercepção em saúde bucal Condições socioeconômicas	5 12 35-44 65-74

No apêndice desse volume apresentamos o detalhamento sobre o plano amostral, faixas etárias e agravos pesquisados.

2 REVISÃO DE LITERATURA

O Instituto Brasileiro de Geografia e Estatística (IBGE) realiza várias pesquisas domiciliares com periodicidades diferentes (mensais, anuais, ou não definidas), com graus variados de abrangência geográfica (nacional, regional, regiões metropolitanas, apenas áreas urbanas, alguns municípios, de capitais, de estados) e complexidade, de acordo não só com os recursos disponíveis mas também com a área temática e objetivos de cada pesquisa. Algumas pesquisas realizadas pelo IBGE tais como a Pesquisa Nacional por Amostra de Domicílios (PNAD, iniciada em 1967); a Pesquisa Mensal de Emprego (PME, desde 1980); a Pesquisa de Orçamentos Familiares; a Pesquisa sobre Padrões de Vida e a Pesquisa de Economia Informal Urbana adotam amostras probabilísticas de domicílios (ALBIERI; BIANCHINI, 1998).

Os desenhos amostrais das pesquisas listadas acima possuem várias semelhanças, as quais incluem amostragem de conglomerados em dois estágios (setores censitários + domicílios) ou mesmo três estágios (municípios + setores censitários + domicílios) com estratificação das unidades primárias de amostragem. As UPA (municípios ou setores censitários) são selecionadas com probabilidade proporcional a uma medida de tamanho – PPT (dentro de cada estrato), da mesma forma que o SB Minas Gerais 2012.

Se, por um lado, a amostra por conglomerados tem como vantagem um menor custo, uma vez que não inclui os gastos de cadastros e na localização de indivíduos, por outro, implica na maior complexidade da análise estatística e, geralmente, um aumento na variância dos estimadores utilizados, acarretando uma diminuição da precisão do estudo. Atualmente, a primeira desvantagem é mais facilmente transposta pelos recursos disponíveis de programação e computação. Porém, em relação ao aumento das variâncias, este fato pode acarretar outras barreiras para o delineamento amostral, principalmente em relação ao tamanho da amostra e à previsão de custos para a realização dos estudos. Isso se deve ao fato de que o aumento da variância dos estimadores é difícil de ser prevista ao se utilizar este tipo de amostragem. O tamanho amostral necessário é proporcional ao efeito do desenho quando fixados a precisão e o intervalo de confiança desejados na amostra por conglomerados. Assim sendo, é importante e útil que se tenha uma noção prévia desse efeito associado à amostragem de conglomerado no planejamento amostral, principalmente, para se definir o tempo e a equipe para o trabalho de campo, bem como para se estipular os custos envolvidos (CORDEIRO, 2001).

Nos países desenvolvidos, os pesquisadores da área de amostragem, já consideram a necessidade de se utilizar programas computacionais específicos para as análises provenientes de bancos de dados com amostras complexas uma vez que levantamentos realizados na área da saúde pública, em geral, adotam delineamento amostral por conglomerados para o sorteio. Porém, muitos pesquisadores ainda utilizam os estimadores da amostragem aleatória simples - AAS para as análises, usando programas estatísticos que não possuem módulos específicos para amostragem com delineamentos complexos o que acarreta análises, em geral, incorretas dos dados (SOUSA; SILVA, 2003).

3 OBJETIVO

O objetivo deste trabalho foi apresentar os aspectos mais importantes do plano amostral do SB Minas Gerais, especialmente no que se refere ao Efeito do Plano Amostral - EPA.

4 METODOLOGIA

Kish (1965 citado por PESSOA; SILVA, 1998) propôs uma medida de efeito do plano amostral sobre a variância de um estimador que denominou efeito do plano amostral. O objetivo desta medida é comparar planos amostrais no estágio de planejamento da pesquisa. O EPA de Kish é uma razão entre variâncias (de aleatorização) de um estimador, calculadas para dois planos amostrais alternativos. Desta forma, a medida EPA tem a seguinte fórmula de cálculo:

$$EPA_{Kish}(\bar{\theta}) = \frac{VVerd(\bar{\theta})}{VAAS(\bar{\theta})},$$

onde:

$\bar{\theta}$ = Estimativa analisada (no nosso caso são a média do ceo-d/CPOD - índice que mede o número de dentes com histórico de cárie - e a proporção de indivíduos com cárie).

$EPA_{Kish}(\bar{\theta})$ = Efeito do Plano Amostral.

$VVerd(\bar{\theta})$ = Variância verdadeira do estimador induzida pelo plano amostral complexo.

$VAAS(\bar{\theta})$ = Variância do estimador induzida pelo plano de amostragem aleatória simples com reposição.

Na prática, o EPA funciona como um “preço” a ser pago pelo pesquisador por ter sua tarefa facilitada ao investigar apenas os *clusters* sorteados, aumentando sua imprecisão devido às possíveis correlações das unidades amostrais dentro e entre os *clusters*. Com ele, pode-se simplificar o cálculo do tamanho amostral usando a amostragem aleatória simples com reposição quando o estudo, na verdade, é feito por conglomeração em vários estágios, corrigindo a subestimação da variância do estimador por meio da aplicação de algum fator de correção baseado no EPA. Uma dificuldade é o pesquisador saber qual o valor a ser atribuído ao EPA, pois, para tal, deveria se ter uma ideia da variabilidade das observações entre e dentro dos *clusters*, o que frequentemente não se dispõe previamente à pesquisa. Normalmente, um EPA de 1,4 ou 1,5, que pode ser traduzido como uma correção no tamanho amostral entre 40% e 50%, segundo alguns autores, parece ser suficiente para resguardar a precisão desejada do pesquisador (LUIZ; MAGNANI, 2000; CORDEIRO, 2001).

No caso do SB Minas Gerais 2012, bem como de outras pesquisas de base populacional, não é possível utilizar a amostragem aleatória simples, pois não se dispõe de listas com todos os indivíduos que podem compor a amostra, isto é, um cadastro de todos indivíduos. Além disso, restrições financeiras impedem este tipo de amostragem. Desta forma, em vários inquéritos, o esquema amostral empregado é a amostragem por conglomerados em vários estágios e vários estudos em saúde bucal vem utilizando o $EPA = 2$ (SILVA; RONCALLI, 2013; PINTO *et al.*, 2018; FRIAS *et al.*, 2016).

Assim, no presente estudo, a partir dos dados do SB Minas Gerais 2012 foram calculados as estimativas pontuais e os intervalos de confiança (IC95%) do ceo-d/CPOD e da proporção de indivíduos com cárie para os três tipos de amostragem, a saber: a amostra aleatória simples com reposição, o plano amostral complexo sem pesos e o plano amostral complexo com pesos. Este cálculo foi realizado para cada domínio definido segundo unidade geográfica e grupos demográficos, ou seja, para os diferentes domínios (Capital, Interior I e Interior II) e para as idades de 5 e 12 anos e as faixas etárias de 15 a 19, 35 a 44, 65 a 74 anos. Em seguida, também foram calculados os erros padrão e os efeitos do plano amostral (EPA) de cada tipo de amostragem. Para estas análises foi utilizado o software R (programação e resultados apresentados no APÊNDICE deste volume).

As tabelas foram então configuradas no programa Excel® versão 2016. Posteriormente foi calculada uma simulação para o $EPA = 2$ utilizado nos estudos epidemiológicos de saúde bucal por meio da expressão 2 vezes a variância do estimador sob o esquema de AASc ($2 \times VAAS(\bar{\theta})$) denominado de (B), onde AASc é a amostra aleatória simples com reposição. Em seguida, foi feita a comparação entre este valor simulado (B) e a variância de fato obtida pela amostra com plano amostral complexo com pesos (A) por meio uma razão apresentada na tabela como Relação (B)/(A) em percentual.

5 RESULTADOS

No SB Minas Gerais 2012 foram realizadas 15 amostras (3 domínios e 5 faixas etárias) na capital Belo Horizonte e nos municípios do Interior I e Interior II. Como o plano amostral desta pesquisa é classificado como complexo, para se extrair informações a partir da análise dos dados, é necessário que se conheça as fases de amostragem apresentadas no apêndice, bem como as particularidades inerentes a estas análises.

A seguir, nas Tabelas 1 e 2 são apresentadas as estimativas de ceo-d/CPOD, proporção de pessoas com cárie e seus respectivos erros padrão e efeito do plano amostral para os domínios do estudo.

Na Tabela 1 estão apresentados os resultados para a proporção de pessoas com cárie por faixa etária e domínio geográfico. Estão destacados os resultados para as prevalências estimadas, erros padrão e efeito do plano amostral para os grupos etários de cinco anos, 12 anos e 65 a 74 anos que, conforme o planejamento amostral, foram as faixas etárias que utilizaram esta variável de proporção para o cálculo da amostra. Para cinco anos, está ilustrada a prevalência de cárie em dentes decíduos, representada pela proporção de indivíduos com $ceo-d \geq 1$ e, aos 12 e 65-74 anos, a prevalência em dentes permanentes (proporção de indivíduos com $CPOD \geq 1$). De modo geral, nota-se baixo impacto do processo de conglomerados e do peso amostral na estimação pontual, mas alto na precisão no erro padrão quando comparadas com as estimativas com e sem pesos.

Analisando a coluna do PAC com pesos da Tabela 1, percebe-se que o único valor do EPA que é menor que 1 é para o domínio Capital na faixa etária de 65-74 anos. Os outros valores são maiores que 1. Os valores dos EPA para os domínios Interior I e Interior II são maiores que os da Capital. Isso pode ser explicado pelo de que, na Capital, o plano amostral teve um estágio de conglomeração a menos. Portanto, para o domínio da Capital, a inflação do EPA parece ter sido excessiva. Para o Interior I, a correção pelo EPA é razoável, com exceção da faixa etária de 12 anos. Já para o domínio Interior II, a correção usando o EPA de 2 foi excessiva, exceto na faixa etária de 5 anos.

Tabela 1 - Estimativas de proporção de indivíduos com cárie, erros padrão e efeito de desenho com e sem ponderação para os grupos etários de acordo com os domínios de estudo. SB Minas Gerais, 2012.

Domínios	Grupos etários	n	Proporção de pessoas com cárie (decíduos/permanentes)										Variância Estimada da AASc mult. pelo EPA=2 (B)	Relação (B)/(A) em %
			AASc com reposição			Com PAC e Sem Pesos			Com PAC e Com Pesos					
			Estimativa (IC95%)	Erro Padrão	EPA	Estimativa (IC95%)	Erro Padrão	EPA	Estimativa (IC95%)	Erro Padrão	Variância (A)	EPA		
Capital	5 anos	200	0,55 (0,48-0,61)	0,035	0,995	0,55 (0,48-0,61)	0,034	0,919	0,55 (0,47-0,62)	0,036	0,00129	1,038	0,00248	192%
	12 anos	262	0,48 (0,42-0,54)	0,031	0,996	0,48 (0,41-0,55)	0,037	1,397	0,48 (0,41-0,56)	0,037	0,00136	1,424	0,00191	140%
	15-19 anos	147	0,68 (0,61-0,76)	0,038	0,993	0,68 (0,62-0,75)	0,033	0,726	0,68 (0,61-0,74)	0,033	0,00106	0,715	0,00290	274%
	35-44 anos	257	0,98 (0,97-1,00)	0,008	0,996	0,98 (0,97-1,00)	0,009	1,341	0,98 (0,96-1,00)	0,010	0,00009	1,391	0,00012	126%
	65-74 anos	246	1,00 (0,99-1,00)	0,004	0,996	1,00 (0,99-1,00)	0,004	0,899	1,00 (0,99-1,00)	0,004	0,00001	0,890	0,00003	226%
Interior I	5 anos	593	0,47 (0,43-0,51)	0,021	0,998	0,47 (0,41-0,53)	0,031	2,234	0,49 (0,43-0,55)	0,032	0,00099	2,354	0,00084	85%
	12 anos	578	0,60 (0,56-0,64)	0,020	0,998	0,60 (0,52-0,67)	0,038	3,493	0,59 (0,52-0,67)	0,037	0,00136	3,264	0,00083	61%
	15-19 anos	587	0,72 (0,68-0,76)	0,019	0,998	0,72 (0,63-0,78)	0,033	3,191	0,72 (0,65-0,78)	0,034	0,00117	3,382	0,00069	59%
	35-44 anos	505	0,97 (0,96-0,99)	0,008	0,998	0,97 (0,95-0,99)	0,010	1,779	0,97 (0,96-0,99)	0,009	0,00007	1,478	0,00011	154%
	65-74 anos	570	0,99 (0,98-1,00)	0,004	0,998	0,99 (0,98-1,00)	0,005	1,371	0,99 (0,98-1,00)	0,005	0,00003	1,659	0,00003	113%
Interior II	5 anos	400	0,53 (0,48-0,58)	0,025	0,998	0,53 (0,47-0,58)	0,028	1,252	0,52 (0,45-0,59)	0,036	0,00127	2,028	0,00125	98%
	12 anos	377	0,67 (0,62-0,72)	0,024	0,998	0,67 (0,60-0,74)	0,034	2,000	0,67 (0,61-0,73)	0,032	0,00104	1,770	0,00118	113%
	15-19 anos	466	0,79 (0,76-0,83)	0,019	0,998	0,79 (0,73-0,85)	0,031	2,682	0,78 (0,70-0,85)	0,038	0,00145	3,898	0,00070	48%
	35-44 anos	442	0,95 (0,93-0,97)	0,010	0,998	0,95 (0,91-0,99)	0,021	3,967	0,95 (0,90-1,00)	0,024	0,00057	5,279	0,00021	37%
	65-74 anos	380	1,00 (0,99-1,00)	0,003	0,998	1,00 (0,99-1,00)	0,003	0,963	1,00 (0,99-1,00)	0,005	0,00002	1,681	0,00001	65%

AASc – amostra aleatória simples com reposição

EPA – Efeito do plano amostral

IC95% - Intervalo de Confiança 95%

Na Tabela 2 as estimativas para médias, erros padrão e EPA do ceo-d/CPOD nos grupos de 15 a 19 anos e de 35 a 44 anos destacados pois foram as faixas etárias que utilizaram este índice para o cálculo da amostra. Analisando a coluna do PAC com pesos, os EPA também podem ser considerados estáveis, porém inflacionados para a maioria dos domínios. Entretanto, os maiores valores para os EPA foram obtidos nos grupos de 15 a 19 nos dois domínios de interior e de 35 a 44 anos para municípios do Interior II, ainda assim não afetando fortemente a estimativa pontual, apenas alargando o intervalo de confiança e aumentando o erro padrão. Esse resultado já era esperado, pois, nesses grupos ocorreram perdas amostrais e, conseqüentemente, diferenças entre os tamanhos das amostras propostos pelo delineamento e aqueles efetivamente alcançados.

Tabela 2 - Estimativas para o ceo-d/CPOD (dentes cariados, perdidos e obturados), erros padrão e efeito de desenho com e sem ponderação para os grupos etários de acordo com os domínios de estudo. SB Minas Gerais, 2012.

Domínios	Grupos etários	n	ceo-d/CPOD										Variância Estimada da AASc mult. pelo EPA=2 (B)	Relação (B)/(A) em %
			AASc com reposição			Com PAC e Sem Pesos			Com PAC e Com Pesos					
			Estimativa (IC95%)	Erro Padrão	EPA	Estimativa (IC95%)	Erro Padrão	EPA	Estimativa (IC95%)	Erro Padrão	Variância (A)	EPA		
Capital	5 anos	200	2,44 (1,97-2,91)	0,240	0,995	2,44 (1,88-3,00)	0,283	1,461	2,42 (1,83-3,01)	0,298	0,08857	1,561	0,11488	130%
	12 anos	262	1,24 (1,01-1,47)	0,119	0,996	1,24 (0,93-1,50)	0,130	1,241	1,26 (0,97-1,55)	0,148	0,02201	1,423	0,02855	130%
	15-19 anos	147	2,34 (1,91-2,78)	0,221	0,993	2,34 (1,88-2,81)	0,236	1,192	2,29 (1,86-2,73)	0,223	0,04979	1,047	0,09763	196%
	35-44 anos	257	16,07 (15,24-16,90)	0,424	0,996	16,07 (15,29-16,84)	0,397	0,919	16,14 (15,33-16,95)	0,415	0,17196	0,949	0,35887	209%
	65-74 anos	246	27,79 (27,03-28,53)	0,382	0,996	27,79 (26,74-28,83)	0,534	2,048	27,56 (26,45-28,67)	0,567	0,32124	2,072	0,29134	91%
Interior I	5 anos	593	2,03 (1,78-2,28)	0,126	0,998	2,03 (1,68-2,38)	0,180	2,048	2,08 (1,74-2,42)	0,174	0,03044	1,899	0,03177	104%
	12 anos	578	2,12 (1,90-2,34)	0,114	0,998	2,12 (1,53-2,70)	0,297	6,798	2,06 (1,50-2,61)	0,283	0,08002	6,403	0,02599	32%
	15-19 anos	587	4,05 (3,71-4,38)	0,171	0,998	4,05 (3,41-4,68)	0,325	3,597	4,11 (3,46-4,76)	0,331	0,10969	3,679	0,05864	53%
	35-44 anos	505	15,85 (15,22-16,49)	0,326	0,998	15,85 (15,01-16,69)	0,427	1,721	15,87 (15,09-16,66)	0,401	0,16066	1,546	0,21209	132%
	65-74 anos	570	29,03 (28,56-29,51)	0,244	0,998	29,03 (28,39-29,67)	0,326	1,785	28,87 (28,06-29,68)	0,412	0,17007	2,734	0,11906	70%
Interior II	5 anos	400	2,32 (2,00-2,63)	0,161	0,998	2,32 (1,97-2,67)	0,179	1,252	2,23 (1,77-2,68)	0,232	0,05388	2,242	0,05204	97%
	12 anos	377	2,64 (2,35-2,93)	0,149	0,998	2,64 (2,10-3,18)	0,275	2,000	2,73 (2,06-3,41)	0,344	0,11847	4,827	0,04426	37%
	15-19 anos	466	4,65 (4,25-5,04)	0,202	0,998	4,65 (3,79-5,51)	0,439	2,682	4,54 (3,62-5,46)	0,469	0,22007	5,441	0,08187	37%
	35-44 anos	442	15,61 (14,77-16,44)	0,427	0,998	15,61 (13,16-18,05)	1,249	3,967	15,29 (12,29-18,30)	1,533	2,35109	12,920	0,36418	15%
	65-74 anos	380	28,96 (28,36-29,55)	0,305	0,998	28,96 (28,11-29,80)	0,429	0,963	28,93 (27,98-29,89)	0,489	0,23885	2,472	0,18660	78%

AASc – amostra aleatória simples com reposição

EPA – Efeito do plano amostral

IC95% - Intervalo de Confiança 95%

Nas tabelas seguintes (Tabelas 3 e 4), nota-se um aumento do efeito do plano amostral, quando comparado às tabelas que apresentam os dados estratificados por idade (Tabelas 1 e 2) mostrando que a estratificação realizada por grupo etário foi necessária, para que as estimativas se aproximassem mais da realidade. Os EPA das Tabelas 3 e 4 do PAC com pesos ficam bem maiores que 2 mostrando que uma correção com um EPA de 2 não seria suficiente caso o plano amostral não considerasse a estratificação por idade. Isto pode ser verificado na coluna da relação de (B)/(A), que é a razão entre o valor da variância simulada com EPA=2 e a variância real encontrada na análise do plano amostral complexo com pesos onde os valores ficaram bem abaixo de 100%, em geral. Este aumento é esperado uma vez que, em termos gerais, quando a população é estratificada, por exemplo, por idade, a variância é menor dentro de um grupo etário do que entre as diversas faixas etárias. Pode-se ainda observar que esta variação para mais no EPA foi maior nos índices ceo-d/CPOD do que na proporção de indivíduos com cárie.

Tabela 3 - Estimativas de proporção de indivíduos com cárie, erros padrão e efeito de desenho com e sem ponderação de acordo com os domínios de estudo. SB Minas Gerais, 2012.

Domínios	n	Proporção de pessoas com cárie (decíduos/permanentes)										Variância Estimada da AASc mult. pelo EPA=2 (B)	Relação (B)/(A) em %
		AAS com reposição			Com PAC e Sem Pesos			Com PAC e Com Pesos					
		Estimativa (IC95%)	Erro Padrão	EPA	Estimativa (IC95%)	Erro Padrão	EPA	Estimativa (IC95%)	Erro Padrão	Variância (A)	EPA		
Capital	1112	0,75 (0,73-0,78)	0,013	0,999	0,75 (0,70-0,80)	0,023	3,267	0,84 (0,80-0,88)	0,019	0,00037	3,102	0,00034	90%
Interior I	2833	0,74 (0,73-0,76)	0,008	1,000	0,74 (0,71-0,78)	0,018	4,604	0,85 (0,82-0,88)	0,015	0,00024	5,264	0,00013	57%
Interior II	2065	0,79 (0,77-0,81)	0,009	1,000	0,79 (0,76-0,82)	0,017	3,616	0,86 (0,83-0,90)	0,017	0,00030	5,136	0,00015	50%

AASc – amostra aleatória simples com reposição

EPA – Efeito do plano amostral

IC95% - Intervalo de Confiança 95%

Para pensar neste efeito em saúde bucal podemos relacionar nos diferentes históricos de doença nas diversas faixas etárias. Na dentição decídua, por exemplo, este perfil é semelhante entre as crianças de mesma idade, mesmo havendo “bolsões de doença”. Quando comparadas as faixas etárias, não podemos verificar tantas semelhanças. Basta pensarmos no perfil de idosos (que são praticamente desdentados) em comparação às crianças de 12 anos (onde atingimos baixos níveis de cárie).

Tabela 4 - Estimativas para o ceo-d/CPOD (dentes cariados, perdidos e obturados), erros padrão e efeito de desenho com e sem ponderação de acordo com os domínios de estudo. SB Minas Gerais, 2012.

Domínios	n	ceo-d/CPOD										Variância Estimada da AASc mult. pelo EPA=2 (B)	Relação (B)/(A) em %
		AAS com reposição			Com PAC e Sem Pesos			Com PAC e Com Pesos					
		Estimativa (IC95%)	Erro Padrão	EPA	Estimativa (IC95%)	Erro Padrão	EPA	Estimativa (IC95%)	Erro Padrão	Variância (A)	EPA		
Capital	1112	10,91 (10,23-11,60)	0,350	0,999	10,91 (8,95-12,88)	1,001	8,212	13,02 (11,19-14,86)	0,938	0,879	7,723	0,245	28%
Interior I	2833	10,36 (9,93-10,79)	0,219	1,000	10,36 (9,00-11,72)	0,694	10,069	13,40 (12,10-14,71)	0,666	0,443	10,202	0,096	22%
Interior II	2065	10,65 (10,15-11,14)	0,253	1,000	10,65 (9,10-12,20)	0,791	9,799	12,82 (11,09-14,54)	0,880	0,774	12,666	0,128	16%

AASc – amostra aleatória simples com reposição

EPA – Efeito do plano amostral

IC95% - Intervalo de Confiança 95%

6 DISCUSSÃO

Análises provenientes de amostras obtidas via planos amostrais complexos tem sido objeto de estudo e discussão na literatura (KORN; GRAUBARD, 1991; SILVA; PESSOA, 2002; SOUSA; SILVA, 2003; SZWARCOWALD; DAMACENA, 2008)

As inferências construídas devem considerar o método traçado para inclusão, por sorteio, de um determinado indivíduo na amostra proveniente do domínio à qual ele pertence (RONCALLI; SILVA, 2013). Isto é, as inferências construídas precisam considerar o plano amostral complexo delineado na fase de planejamento do estudo. No SB Minas Gerais 2012, o modelo geral usado foi o da amostragem por conglomerados em múltiplos estágios, nos quais as unidades amostrais foram sorteadas com probabilidade proporcional ao número de domicílios existentes em cada uma.

Dito de outra forma, ao se ignorar o plano amostral, a análise estatística tradicional, utilizando a suposição de amostra aleatória simples, pode produzir incorreções, comprometendo os resultados da estimativa pontual, o erro padrão, os intervalos de confiança, os testes de hipóteses, e as conclusões da pesquisa. Desta forma, ao utilizar análises em amostras por conglomerados, é necessário que haja correções para produzirem inferências estatisticamente válidas (QUEIROZ *et al.*, 2009)

Os EPA apresentados consolidam os resultados obtidos para médias e proporções nos três domínios geográficos e nos 5 domínios etários com algumas exceções onde houve maiores discrepâncias, por exemplo, na faixa etária de adultos. Ao que parece, a característica conservadora do delineamento com cálculo amostral utilizando $EPA = 2$ ao invés de $EPA=1,4$ ou $1,5$ preservou os critérios de precisão e o efeito de conglomerados, mantendo-os em níveis aceitáveis para algumas faixas etárias, porém em outras isto não ocorreu.

É sabido que neste inquérito a faixa etária de 35-44 foi a que obteve menor taxa de resposta (média dos 3 domínios de 62,6%), justificando os valores discrepantes do EPA encontrado (MINAS GERAIS, 2013). Em inquéritos deste tipo deve-se levar em consideração a necessidade de se persistir na busca pelos ausentes, atuando em dias e horários diversificados, especialmente para esta faixa etária por esta se apresentar como população economicamente ativa e serem mais difíceis de serem encontrados nos domicílios em horário comercial.

Apesar de, na Capital, também terem ocorrido baixas taxas de resposta em algumas faixas etárias (15-19 anos e 35-44 anos), como é possível verificar no anexo deste trabalho, isto parece não ter afetado o efeito do plano amostral uma vez que as amostras na capital possuíam um número maior de indivíduos e similaridades entre estes maiores que nos domínios de interior, que eram compostos de diversos municípios.

Como visto na seção de resultados, o efeito da estratificação por idade também foi positivo para a convergência dos dados para mais próximo dos dados reais da população, uma vez que a variância foi menor nas tabelas de resultados onde a estratificação estava presente do que naquelas em que não estava. Isso se deve ao fato de que os problemas bucais em determinada faixa etária são similares pelo contexto político, econômico e social em que estes indivíduos estiveram inseridos ao longo de sua história de vida. Por exemplo, podemos citar a faixa etária de idosos, que apresentam grandes perdas dentárias, fato diretamente relacionado à ausência de políticas públicas de saúde bucal na maior parte da sua vida, tais como direito ao acesso, ausência de fluoretação das águas e de pasta de dente com flúor que foram objeto de regulamentação mais recente.

7 CONSIDERAÇÕES FINAIS

Os resultados apresentados devem ser levados em consideração para os planos amostrais de inquéritos epidemiológicos de saúde bucal futuros caso se utilizem os mesmos seguimentos demográficos, especialmente se sua operacionalização se der por equipes de profissionais dos serviços locais de saúde para a realização das entrevistas e exames com amostras domiciliares. Assim sendo, é de extrema importância que as análises estatísticas realizadas com bancos de dados provenientes de amostras complexas levem em consideração o delineamento amostral realizado. É necessário ressaltar que o SB Minas Gerais 2012, assim como o SB Brasil 2010, são modelos de levantamentos epidemiológicos pensados para serem utilizados pelos serviços de saúde bucal em suas práticas cotidianas, como ferramentas para as ações de planejamento e avaliação das ações e serviços de saúde. Assim, é necessário que as amostras aliem a questão operacional com a sua representatividade. Além disso, em estudos epidemiológicos futuros, ao se planejar a amostragem, deve ser levada em consideração a necessidade de variar o EPA de acordo com as faixas etárias.

REFERÊNCIAS

ALBIERI, S.; BIANCHINI, Z.M. Uma revisão dos principais aspectos dos planos amostrais das pesquisas domiciliares realizadas pelo IBGE. Rio de Janeiro: IBGE, Departamento de Metodologia, 1998. 24p. - (Textos para discussão, ISSN 0103-6661; n. 91).

BELL, B. A.; ONWUEGBUZIE, A. J.; FERRON, J. M.; JIAO, Q. G.; HIBBARD, S. T.; KROMREY, J. D. Use of design effects and sample weights in complex health survey data: a review of published articles using data from 3 commonly used adolescent health surveys. **American Journal of Public Health**, v. 102, n. 7, p. 1399-1405, 2012

BRASIL. Ministério da Saúde. Projeto SB Brasil 2010: Pesquisa nacional de saúde bucal - resultados principais. Brasília: Ministério da Saúde; 2011. 92 p. il.

CORDEIRO, R. Efeito do desenho em amostragem de conglomerado para estimar a distribuição de ocupações entre trabalhadores. **Rev. Saúde Pública**, São Paulo, v. 35, n. 1, p. 10-15, Fev. 2001. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102001000100002&lng=en&nrm=iso>. Acesso em: 10 Abril 2019. <http://dx.doi.org/10.1590/S0034-89102001000100002>.

CUNNINGHAM, T. D.; JOHNSON, R. E. Design effects for sample size computation in three-level designs. **Statistical Methods in Medical Research**, Londres, v.25, n. 2, p. 505-519, 2012. Acesso em: 10 Agosto 2019. Disponível em: <<http://dx.doi.org/10.1177/0962280212460443>>.

FRIAS, A. C.; PEREIRA, A. C.; VIEIRA, V. **Pesquisa estadual de saúde bucal: relatório final**. Águas de São Pedro: Livronovo, 2016. Acesso em: 18 Agosto 2019. Disponível em: <http://www.saude.sp.gov.br/resources/ses/perfil/profissional-da-saude/areas-tecnicas-da-ses/e_book_relatorio_sb_sp_2015.pdf>

GLENN, B. A.; BASTANI, R; MAXWELL, A. E. The perils of ignoring design effects in experimental studies: Lessons from a mammography screening trial. **Psychology & Health**, v. 28, n. 5, p, 593–602, 2013 Acesso em: 18 Agosto 2019. Disponível em: <<http://dx.doi.org/10.1080/08870446.2012.756880>>

GOUDA, H. N.; POWLES, J. W. The science of epidemiology and the methods needed for public health assessments: a review of epidemiology textbooks. **BMC Public Health**, v. 14, n. 139, 2014. Acesso em: 18 Agosto 2019. Disponível em: < <http://dx.doi.org/10.1186/1471-2458-14-139>>

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Pesquisa nacional por amostra de domicílios. Rio de Janeiro; 2016. Acesso em: 18 Agosto 2019. Disponível em: < <https://ww2.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2015/default.shtml>>

KORN, E.L.; GRAUBARD, B.I. Epidemiologic Studies Utilizing Surveys: Accounting for the Sampling Design. **American Journal of Public Health**. 1991, V. 81, n. 9, Set. 1991.

LISA, G. J.; CHEN, Y. H.; SILVA-SANTISTEBAN, A.; RAYMOND, H. F. An empirical examination of respondent driven sampling design effects among HIV risk groups from studies

conducted around the world. **AIDS and Behavior**, v. 17, p2202-2210, 2013. Disponível em: <<http://dx.doi.org/10.1007/s10461-012-0394-8>>

LUIZ, R. R.; MAGNANINI, M. M. F. A lógica da determinação do tamanho da amostra em investigações epidemiológicas. **Cadernos de Saúde Coletiva**, Rio de Janeiro, v. 8, n. 2, p. 9-28, ago./dez. 2000.

MACHADO, E.N.M. *et al.* Fator de alocação de recursos financeiros para atenção à saúde. Belo Horizonte: Fundação João Pinheiro/Secretaria de Estado da Saúde, 2003 (mimeo).

MINAS GERAIS. Secretaria de Estado de Saúde. Deliberação CIB-SUS/MG nº 1.040, de 14 de fevereiro de 2012. Institui o Projeto SB Minas Gerais – Pesquisa das condições de saúde bucal da população mineira. Minas Gerais. 16 fev 2012; Caderno1:12.

MINAS GERAIS. Secretaria de Estado de Saúde. Metodologia de alocação equitativa de recursos: uma proposta para Minas Gerais. / Mônica Viegas Andrade et al. Belo Horizonte: 2004. 63p. il.

MINAS GERAIS. Secretaria de Estado de Saúde. Diretoria de Saúde Bucal. SB Minas Gerais: Pesquisa das Condições de Saúde Bucal da População Mineira – Resultados Principais. Belo Horizonte: Editora Autêntica; 2013a. 80 p.

PESSOA, D.G.C; SILVA, P.L.N. **Análise de Dados Amostrais Complexos**. Março, 1998. Disponível em: <<http://www.ie.ufrj.br/download/livro.pdf>>

PINTO, R.S. **Uso de serviços públicos de saúde bucal pela população adulta**. 2015. 115f. Dissertação (Mestrado em Odontologia) – Faculdade de Odontologia, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

PINTO, R. S.; LEAL, D. L.; SANTOS, J. S.; RONCALLI, A. G. Projeto SB Minas Gerais 2012: Pesquisa das Condições de Saúde Bucal da População Mineira – Métodos e Resultados Principais. **Arquivos em Odontologia**, Belo Horizonte, 54: e14, 2018. Acesso em: 10 Abril 2019. Disponível em: <<https://periodicos.ufmg.br/index.php/arquivesodontologia/article/view/3733>>

QUEIROZ, R.C.S.; PORTELA, M.C.; VASCONCELLOS, M.T.L. Pesquisa sobre as condições de saúde bucal da população brasileira (SB Brasil 2003): seus dados não produzem estimativas populacionais, mas há possibilidade de correção. **Cad. Saude Publica**, Rio de Janeiro, v. 25, n. 1, Jan. 2009. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102311X2009000100005&lng=en&nrm=iso>. Acesso em: 15 Abril 2019.

SILVA, N. N.; RONCALLI, A. G. Plano amostral, ponderação e efeitos do delineamento da Pesquisa Nacional de Saúde Bucal. **Rev. Saúde Pública**, São Paulo, v.47, supl. 3, p. 3-11, Dez. 2013. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102013000900003&lng=en&nrm=iso>. Acesso em 10 Jul. 2019. <http://dx.doi.org/10.1590/S0034-8910.2013047004362>.

SILVA, P. L. N., PESSOA, D. G. C. P., LILA, M. F. Análise estatística de dados da PNAD: incorporando a estrutura do plano amostral. **Ciência & Saúde Coletiva** 2002; 7(4): 659-670

SOUSA, M.H.; SILVA, N.N. Estimativas obtidas de um levantamento complexo. **Rev. Saúde Pública**, São Paulo, v. 37, n. 5, p. 662-670, Out. 2003. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102003000500018&lng=en&nrm=iso>. Acesso em 10 Jul. 2019. <http://dx.doi.org/10.1590/S0034-89102003000500018>.

SZWARCWALD, C.L; DAMACENA, G.N. Amostras complexas em inquéritos populacionais: planejamento e implicações na análise estatística dos dados. **Rev. bras. epidemiol.**, São Paulo, 2010. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-790X2008000500004&lng=en&nrm=iso>.

WORLD HEALTH ORGANIZATION. Oral health surveys: basic methods. 4 ed. Geneva: ORH/EPID, 1997.

ANEXO 1 – Detalhamento do plano amostral da do Projeto SB Minas Gerais – Pesquisa das Condições de Saúde Bucal da População Mineira¹

1 Área e População de Estudo - SB Minas Gerais 2012

O presente estudo utilizará o banco de dados do Projeto SB Minas Gerais – Pesquisa das Condições de Saúde Bucal da População Mineira (MINAS GERAIS, 2013), um inquérito epidemiológico de abrangência estadual, voltado para a investigação de aspectos relacionados à saúde bucal, tais como os principais agravos de saúde e utilização de serviços odontológicos, entre outros.

2 Características do Projeto SB Minas Gerais - Pesquisa das condições de saúde bucal da população mineira

Trata-se de uma pesquisa de base estadual, com representatividade para o estado de Minas Gerais, com 2 agrupamentos de municípios de acordo com os fatores de alocação.

A população em estudo foi resultante de um processo de amostragem probabilística por conglomerados, considerando-se os grupos etários e o fator de alocação dos municípios, índice produzido pela Fundação João Pinheiro (MACHADO *et al.*, 2003; MINAS GERAIS, 2004).

O fator de alocação foi construído a partir da associação de dois índices:

-Índice de Necessidade em Saúde: indicador composto por um conjunto de seis variáveis epidemiológicas e socioeconômicas;

-Índice de Porte econômico: corresponde ao valor per capita do ICMS de cada município, trabalhado por uma expressão logarítmica que expressa a capacidade do município financiar com recursos próprios, os cuidados com a saúde dos cidadãos.

¹ Texto extraído de PINTO, R.S. **Uso de serviços públicos de saúde bucal pela população adulta**. 2015. 115f. Dissertação (Mestrado em Odontologia) – Faculdade de Odontologia, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

Para a Secretaria de Estado de Saúde, os municípios foram classificados em quartis, a saber:

Quadro 2 - Classificação do Fator de Alocação segundo faixas.

Grupo	Fator de Alocação
1 (1ª faixa)	1,0805 a 1,3364
2 (2ª faixa)	1,3365 a 1,4392
3 (3ª faixa)	1,4393 a 1,5893
4 (4ª faixa)	1,5894 a 2

Os municípios classificados no grupo 1 são aqueles que tem menor necessidade relativa de recursos financeiros. Os do grupo 4 são os que tem maior necessidade. A Figura 1 mostra a distribuição dos municípios de Minas Gerais segundo grupo (fator de alocação):

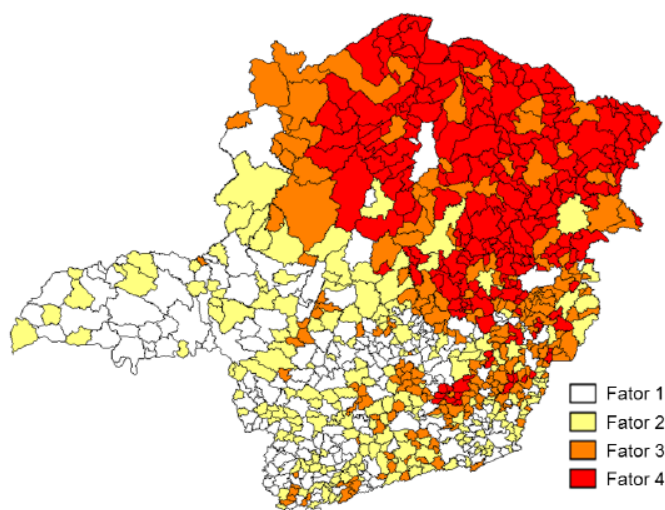


Figura 1 - Distribuição dos municípios mineiros segundo quartis (fator de alocação)

Os municípios foram agrupados como dois domínios onde as amostras são representativas de cada domínio, bem como do estado de Minas Gerais. A capital Belo Horizonte não foi incluída no sorteio uma vez que o Projeto SB Brasil 2010 contemplou amostra representativa da cidade. Os dados de 2010 de Belo Horizonte foi agregado ao banco com os dados do Interior. O primeiro domínio foi composto de uma amostra de 30 municípios com fator de alocação 1 ou 2 e o segundo domínio composto também de 30 municípios com fator de alocação 3 e 4 conforme Figura 2.

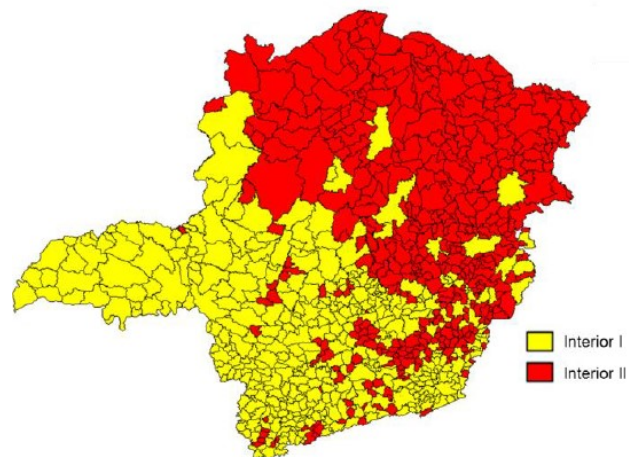


Figura 2 - Distribuição dos municípios mineiros para composição dos domínios da pesquisa.

Na Figura 3 é possível ver a distribuição espacial dos municípios sorteados:

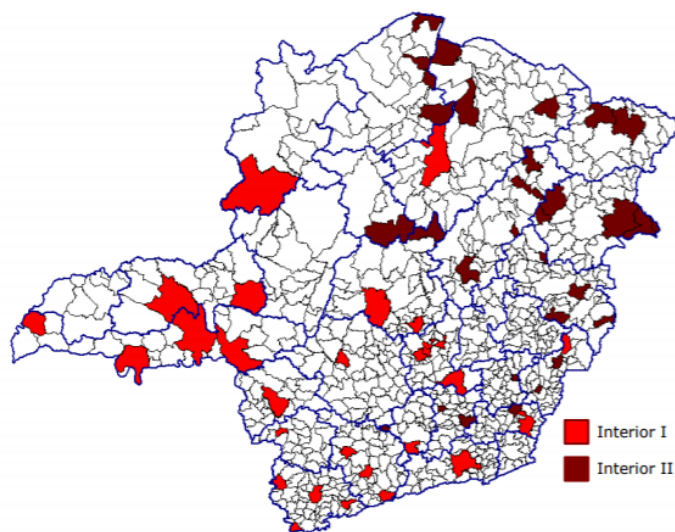


Figura 3 - Municípios sorteados de acordo com o domínio.

3 Plano Amostral

O Projeto SB Brasil 2010 (BRASIL, 2010) foi considerado uma importante experiência na Epidemiologia em Saúde Bucal no Brasil, entre outros aspectos, pelo fato de ter experimentado uma metodologia em seu desenho amostral bastante inovadora para a área de saúde bucal. A partir da experiência acumulada em pesquisas anteriores, foi possível chegar a um delineamento que agregou eficiência e factibilidade, gerando um conjunto de dados extremamente valioso para o conhecimento da realidade epidemiológica brasileira de saúde bucal. Desse modo, a proposta foi que o Plano Amostral do SB Minas Gerais mantivesse a mesma lógica do SB Brasil 2010, obviamente fazendo as devidas adaptações para a realidade estadual.

3.1 Domínios e Unidades Primárias de Amostragem (UPA)

As UPA, representadas pelos municípios, onde estes foram classificados em quatro categorias pelo Fator de Alocação, um indicador criado pela Fundação João Pinheiro e utilizado pela Secretaria de Estado de Saúde de Minas Gerais para a alocação equitativa de recursos (MACHADO *et al.*, 2003; MINAS GERAIS, 2004).

Considerando a viabilidade do estudo, foram estabelecidos dois domínios, representados pelos municípios classificados como “1” ou “2” e por aqueles classificados como “3” ou “4”. Em cada domínio, foram sorteados 30 municípios titulares e dois reservas, para o caso de desistência ou perda de UPA por questões operacionais. Ao todo, portanto, 60 municípios deverão compor a amostra do estudo.

3.2 Unidades Secundárias de Amostragem (USA): Setores Censitários

A base de dados dos setores censitários do Censo 2010 para o estado de Minas Gerais foi obtida no sítio do IBGE (www.ibge.gov.br). Em seguida, o banco de dados foi preparado a partir do ajuste das variáveis, seguindo o fluxograma ilustrado na Figura 4 a seguir.

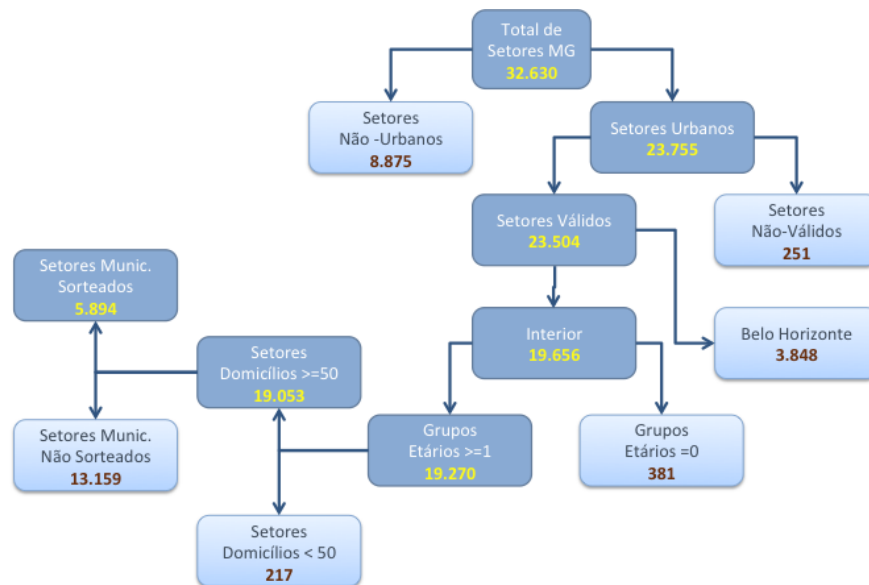


Figura 4 - Fluxograma para ajuste do banco de dados dos setores censitários.

Os 5.894 setores remanescentes foram então separados por municípios e, em cada um deles, foi realizado o sorteio de 4 setores pela técnica PPT (probabilidade proporcional ao tamanho), semelhante ao sorteio dos municípios, sendo que, neste caso, o número de domicílios foi usado como referência.

3.3 Idades-índices e grupos etários do Projeto SB Minas Gerais

A Organização Mundial da Saúde (OMS) sugere, para estudos de saúde bucal, a composição da amostra em determinadas idades-índice e grupos etários os quais serão utilizados na presente pesquisa com algumas modificações. As descrições colocadas a seguir foram retiradas parcialmente da 4ª edição do Manual da OMS, de 1997 (WHO, 1997).

5 anos: Esta idade é de interesse em relação aos níveis de doenças bucais na dentição decídua, uma vez que podem exibir mudanças em um período de tempo menor que a dentição permanente em outras idades-índice, além de ser usada internacionalmente para aferição do ataque de cárie em dentes decíduos.

12 anos: Esta idade é especialmente importante, pois foi escolhida como a idade de monitoramento global da cárie para comparações internacionais e o acompanhamento das tendências da doença.

15 a 19 anos: Considerando a possibilidade de comparação com os dados de 1986 e levando-se em conta, ainda, que, ao se trabalhar com idades restritas como 15 e 18 anos dificulta-se bastante o delineamento amostral (em função da sua proporção no conjunto da população), foi definido manter a faixa etária de 15 a 19 anos. Caso se deseje uma análise mais apurada de cada idade em particular, pertencente a este intervalo, os dados poderão ser agregados por porte ou região ou ainda outra variável que permita este nível de análise.

35 a 44 anos: Este grupo etário é o grupo padrão para avaliação das condições de saúde bucal em adultos. O efeito total da cárie dentária, o nível de gravidade do envolvimento periodontal e os efeitos gerais do tratamento prestado podem ser monitorados usando-se dados deste grupo etário.

65 a 74 anos: Este grupo etário tem se tornado mais importante com as mudanças na distribuição etária e no aumento da expectativa de vida que vem ocorrendo em muitos países. Os dados deste grupo são necessários tanto para o planejamento adequado do tratamento para os mais idosos como para o monitoramento dos efeitos gerais dos serviços odontológicos prestados a uma população.

Os indivíduos de cada grupo etário e idade-índice foram avaliados com relação às doenças bucais explicitadas anteriormente e de acordo com o

Quadro 3 a seguir:

Quadro 3 - Condições e idades/grupos etários a serem pesquisados.

Idade / Grupo Etário (anos)	Cárie Dentária		Traumatismo Dentário	Condição Periodontal		Fluorose	Condição de oclusão dentária		Edentulismo
	Coroa	Raiz		CPI	PIP		OMS (1987)	DAI	
5	●						●		
12	●		●	●		●		●	
15 a 19	●			●				●	●
35 a 44	●	●		●	●				●
65 a 74	●	●		●	●				●

Segundo Pinto *et al.* (2018):

“para cada grupo etário e cada domínio de Interior a prevalência de cárie e a média do CPOD foram utilizadas como referência para o cálculo do tamanho da amostra associado a uma determinada margem de erro. Para as idades de 5 e 12 anos e para o grupo etário de 65 a 74 anos, adotou-se o coeficiente de variação como indicador de precisão para estimativas de prevalências. Escolheu-se o número mínimo de exames, esperando-se que as prevalências estimadas (P) fossem maiores que 10% e que seus erros padrão $[ep(p)]$ não ultrapassassem 15% desses valores. Nos grupos de 15 a 19 e de 35 a 44 anos, calculou-se o tamanho (n) da amostra final pela expressão $n = [(sx \cdot 1,96)/m]^2$, onde 1,96 é o termo da distribuição normal correspondente ao intervalo de confiança de 95%, e “m” é a margem tolerada para o erro inerente ao processo de amostragem aleatória simples. Estimativas para (sx) foram calculadas com dados do SBBrasil 2010, tendo como referência o domínio relativo aos municípios do interior da Região Sudeste. Os resultados iniciais foram corrigidos para compensar o efeito de taxas de respostas em torno de 80% e efeito de desenho (deff) igual a 2 para proteger o impacto do delineamento por conglomerados sobre a precisão inicialmente fixada, admitindo o processo de amostragem como aleatória simples.”

O relatório final está disponível no *site* da Secretaria de Estado de Saúde de Minas Gerais em <https://bit.ly/2kypUPC> e o banco de dados pode ser solicitado enviando o formulário disponível em <https://bit.ly/30y9h67> para o e-mail saudebucal@saude.mg.gov.br.

ANEXO 2 - Número de domicílios e indivíduos pesquisados e respectivas taxas de resposta de acordo com domínio e grupo etário.²

Grupo Etário	Domínio	Domicílios			Indivíduos		
		DS	DNP	DP	ER	ENR	TR
5 anos	Capital	250	50	200	200	0	80,0
	Interior I	500	0	593	588	5	117,6
	Interior II	500	100	400	392	8	78,4
	Estado	1.250	57	1.193	1.180	13	94,4
12 anos	Capital	250	0	262	262	0	104,8
	Interior I	500	0	578	576	2	115,2
	Interior II	500	123	377	371	6	74,2
	Estado	1.250	33	1.217	1.209	8	96,7
15 a 19 anos	Capital	200	51	149	147	2	73,5
	Interior I	750	163	587	580	7	77,3
	Interior II	750	284	466	462	4	61,6
	Estado	1.700	498	1.202	1.189	13	69,9
35 a 44 anos	Capital	487	227	260	257	3	52,8
	Interior I	700	195	505	494	11	70,6
	Interior II	700	258	442	431	11	61,6
	Estado	1.887	680	1.207	1.182	25	62,6
65 a 74 anos	Capital	250	3	247	246	1	98,4
	Interior I	500	0	570	566	4	113,2
	Interior II	500	120	380	380	0	76,0
	Estado	1.250	53	1.197	1.192	5	95,4
Total	Capital	1.437	319	1.118	1.112	6	77,4
	Interior I	2.950	117	2.833	2.804	29	95,1
	Interior II	2.950	885	2.065	2.036	29	69,0
	Estado	7.337	1.321	6.016	5.952	64	81,1

Legenda: DS=Domicílios Sorteados; DNP=Domicílios Não Pesquisados (Fechados e Recusas); DP=Domicílios Pesquisados; ER=Exames Realizados; ENR=Exames Não Realizados (Não Autorizado, Não Permitido e Ausência do Morador); TR = Taxa de Resposta, expressa em percentual (DP/DS x ER/DP).

² Extraído de MINAS GERAIS. Secretaria de Estado de Saúde. Diretoria de Saúde Bucal. SB Minas Gerais: Pesquisa das Condições de Saúde Bucal da População Mineira – Resultados Principais. Belo Horizonte: Editora Autêntica; 2013a. 80 p.

APÊNDICE – Programação e resultados das análises no R.

1 Capital

1.1 Com plano amostral complexo e com pesos

```

> #Mean:
> #svyby(~CEO,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T)      #without
replacement
> #confint(svyby(~CEO,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T))
> #svyby(~CPOD,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T)      #without
replacement
> #confint(svyby(~CPOD,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T))
> svymean(~CEOCPOD, design=dcluslsub, deff="replace",na.rm=T)
      mean      SE  DEff
CEOCPOD 13.0237  0.9375  7.7225
> confint(svymean(~CEOCPOD,design=dcluslsub,na.rm=T))
      2.5 %   97.5 %
CEOCPOD 11.18619 14.86112
> svyby(~CEOCPOD,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T)      #without
replacement
      GR_ETARIO  CEOCPOD      se DEff.CEOCPOD
1          1  2.422973 0.2976022  1.5614265
2          2  1.263649 0.1483679  1.4228146
3          3  2.294750 0.2231303  1.0468790
4          4 16.139354 0.4146789  0.9494996
5          5 27.562443 0.5667806  2.0719241
> confint(svyby(~CEOCPOD,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T))
      2.5 %   97.5 %
1  1.8396833  3.006263
2  0.9728529  1.554444
3  1.8574228  2.732078
4 15.3265981 16.952110
5 26.4515738 28.673313
> # svyby(~CEOCPOD,~GR_ETARIO, design=dcluslsub,svymean, deff="replace",na.rm=T)
#without replacement
> # confint(svyby(~CEOCPOD,~GR_ETARIO, design=dcluslsub,svymean, deff="replace",na.rm=T))
>
> #Prop:
> svymean(~PCAR_SUM, design=dcluslsub, deff="replace",na.rm=T)
      mean      SE  DEff
PCAR_SUM 0.841311 0.019254 3.1018

```

```

> confint(svymean(~PCAR_SUM, design=dclus1sub, na.rm=T))
      2.5 %    97.5 %
PCAR_SUM 0.8035736 0.8790491
> # svyby(~PCAR_SUM, ~GR_ETARIO, design=dclus1sub, svymean, deff="TRUE", na.rm=T) #without
replacement
> # confint(svyby(~PCAR_SUM, ~GR_ETARIO, design=dclus1sub, svymean, deff="TRUE", na.rm=T))
> svyby(~PCAR_SUM, ~GR_ETARIO, design=dclus1sub, svymean, deff="replace", na.rm=T)
#without replacement
      GR_ETARIO  PCAR_SUM          se DEff.PCAR_SUM
1          1  0.5454143  0.035967775    1.0383360
2          2  0.4834268  0.036913254    1.4241068
3          3  0.6757042  0.032546570    0.7154417
4          4  0.9824795  0.009615598    1.3911805
5          5  0.9959873  0.003803510    0.8904619
> confint(svyby(~PCAR_SUM, ~GR_ETARIO, design=dclus1sub, svymean, deff="replace", na.rm=T))
      2.5 %    97.5 %
1 0.4749187 0.6159098
2 0.4110782 0.5557755
3 0.6119141 0.7394943
4 0.9636333 1.0013258
5 0.9885326 1.0034421

```

1.2 Com plano amostral complexo e sem pesos

```

> svymean(~CEOCPOD, design=dcluslsub, deff="replace",na.rm=T)

      mean      SE  DEff
CEOCPOD 10.9141  1.0011  8.2118

> confint(svymean(~CEOCPOD,design=dcluslsub,na.rm=T))

      2.5 %   97.5 %
CEOCPOD 8.952022 12.87624

> svyby(~CEOCPOD,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T) #without
replacement

      GR_ETARIO  CEOCPOD      se DEff.CEOCPOD
1           1  2.440000 0.2833065   1.4606085
2           2  1.240458 0.1301290   1.2413129
3           3  2.342282 0.2361171   1.1918123
4           4 16.065385 0.3970007   0.9192253
5           5 27.785425 0.5339583   2.0479069

> confint(svyby(~CEOCPOD,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T))

      2.5 %   97.5 %
1  1.8847294  2.995271
2  0.9854099  1.495506
3  1.8795010  2.805063
4 15.2872776 16.843492
5 26.7388861 28.831964

> # svyby(~CEOCPOD,~GR_ETARIO, design=dcluslsub,svymean, deff="replace",na.rm=T)
#without replacement

> # confint(svyby(~CEOCPOD,~GR_ETARIO, design=dcluslsub,svymean, deff="replace",na.rm=T))
>
> #Prop:
> svymean(~PCAR_SUM, design=dcluslsub, deff="replace",na.rm=T)

      mean      SE  DEff
PCAR_SUM 0.750447 0.023404 3.267

> confint(svymean(~PCAR_SUM,design=dcluslsub,na.rm=T))

      2.5 %   97.5 %
PCAR_SUM 0.7045763 0.7963182

> # svyby(~PCAR_SUM,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T) #without
replacement

> # confint(svyby(~PCAR_SUM,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T))

> svyby(~PCAR_SUM,~GR_ETARIO, design=dcluslsub,svymean, deff="replace",na.rm=T)
#without replacement

      GR_ETARIO  PCAR_SUM      se DEff.PCAR_SUM
1           1  0.5450000 0.033833844   0.9186453

```

```
2      2 0.4809160 0.036548732    1.3966200
3      3 0.6845638 0.032556022    0.7264385
4      4 0.9846154 0.008856689    1.3411868
5      5 0.9959514 0.003837854    0.8986091

> confint(svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub, svymean, deff="replace", na.rm=T))
      2.5 %    97.5 %
1 0.4786869 0.6113131
2 0.4092818 0.5525502
3 0.6207551 0.7483724
4 0.9672566 1.0019742
5 0.9884294 1.0034735
```


1.3 Sem peso, sem estrato e cluster (Amostra Aleatória Simples com reposição)

```

> svymean(~CEOCPOD, design=dclus1sub, deff="replace",na.rm=T)

      mean      SE  DEff
CEOCPOD 10.91413  0.34922 0.9993

> confint(svymean(~CEOCPOD,design=dclus1sub,na.rm=T))

      2.5 %   97.5 %
CEOCPOD 10.22968 11.59859

> svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T) #without
replacement

      GR_ETARIO  CEOCPOD      se DEff.CEOCPOD
1           1  2.440000 0.2396697          Inf
2           2  1.240458 0.1194854          Inf
3           3  2.342282 0.2209393          Inf
4           4 16.065385 0.4235994          Inf
5           5 27.785425 0.3816658          Inf

> confint(svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))

      2.5 %   97.5 %
1  1.970256  2.909744
2  1.006271  1.474645
3  1.909249  2.775315
4 15.235145 16.895624
5 27.037374 28.533476

> # svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T)
#without replacement

> # confint(svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T))
>
> #Prop:
> svymean(~PCAR_SUM, design=dclus1sub, deff="replace",na.rm=T)

      mean      SE  DEff
PCAR_SUM 0.750447 0.012944 0.9993

> confint(svymean(~PCAR_SUM,design=dclus1sub,na.rm=T))

      2.5 %   97.5 %
PCAR_SUM 0.7250782 0.7758163

> # svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T) #without
replacement

> # confint(svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))
> svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T)
#without replacement

      GR_ETARIO  PCAR_SUM      se DEff.PCAR_SUM
1           1  0.5450000 0.035214786    0.9951654

```

```
2      2 0.4809160 0.030870161    0.9963488
3      3 0.6845638 0.038072002    0.9934537
4      4 0.9846154 0.007633541    0.9963195
5      5 0.9959514 0.004040715    0.9961170

> confint(svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub, svymean, deff="replace", na.rm=T))
      2.5 %    97.5 %
1 0.4759803 0.6140197
2 0.4204116 0.5414204
3 0.6099440 0.7591835
4 0.9696539 0.9995769
5 0.9880318 1.0038711
```

2 Interior I

2.1 Com plano amostral complexo e com pesos

```

> > # 5894 setores eligiveis

> #Mean:

> #svyby(~CEO,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T)      #without
replacement

> #confint(svyby(~CEO,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))

> #svyby(~CPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T)      #without
replacement

> #confint(svyby(~CPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))

> svymean(~CEOCPOD, design=dclus1sub, deff="replace",na.rm=T)

      mean      SE  DEff
CEOCPOD 13.40211  0.66577 10.202

> confint(svymean(~CEOCPOD,design=dclus1sub,na.rm=T))

      2.5 %   97.5 %
CEOCPOD 12.09723 14.70699

> svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T)      #without
replacement

      GR_ETARIO  CEOCPOD      se DEff.CEOCPOD
1           1  2.082512 0.1744604    1.899290
2           2  2.059044 0.2828803    6.403378
3           3  4.107846 0.3311908    3.679275
4           4 15.873469 0.4008199    1.545860
5           5 28.870229 0.4123941    2.734478

> confint(svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))

      2.5 %   97.5 %
1  1.740576  2.424448
2  1.504609  2.613479
3  3.458724  4.756968
4 15.087876 16.659062
5 28.061951 29.678506

> # svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T)
#without replacement

> # confint(svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T))
>
> #Prop:

> svymean(~PCAR_SUM, design=dclus1sub, deff="replace",na.rm=T)

      mean      SE  DEff
PCAR_SUM 0.851105 0.015347 5.2635

> confint(svymean(~PCAR_SUM,design=dclus1sub,na.rm=T))

```

```

                2.5 %    97.5 %
PCAR_SUM 0.8210257 0.8811846
> # svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T)      #without
replacement
> # confint(svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))
> svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T)
#without replacement

GR_ETARIO  PCAR_SUM          se DEff.PCAR_SUM
1          1 0.4867883 0.031516987          2.353826
2          2 0.5939281 0.036937164          3.264120
3          3 0.7157025 0.034270060          3.382375
4          4 0.9739854 0.008623639          1.479251
5          5 0.9906197 0.005204892          1.658868
> confint(svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T))
                2.5 %    97.5 %
1 0.4250162 0.5485605
2 0.5215326 0.6663237
3 0.6485345 0.7828706
4 0.9570834 0.9908874
5 0.9804183 1.0008211

```

2.2 Com plano amostral complexo e sem pesos

```

> #Mean:
> #svyby(~CEO,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T)      #without
replacement
> #confint(svyby(~CEO,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))
> #svyby(~CPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T)      #without
replacement
> #confint(svyby(~CPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))
> svymean(~CEOCPOD, design=dclus1sub, deff="replace",na.rm=T)

      mean      SE  DEff
CEOCPOD 10.36357  0.69436 10.069
> confint(svymean(~CEOCPOD,design=dclus1sub,na.rm=T))

      2.5 %   97.5 %
CEOCPOD 9.00266 11.72448
> svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T)      #without
replacement

      GR_ETARIO  CEOCPOD      se DEff.CEOCPOD
1          1  2.028668 0.1802185    2.047687
2          2  2.121107 0.2970551    6.798443
3          3  4.047700 0.3245169    3.596951
4          4 15.853465 0.4270226    1.721489
5          5 29.033333 0.3257546    1.784935
> confint(svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))

      2.5 %   97.5 %
1  1.675446  2.381889
2  1.538890  2.703324
3  3.411659  4.683742
4 15.016517 16.690414
5 28.394866 29.671801
> # svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T)
#without replacement
> # confint(svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T))
>
> #Prop:
> svymean(~PCAR_SUM, design=dclus1sub, deff="replace",na.rm=T)

      mean      SE  DEff
PCAR_SUM 0.742676 0.017626 4.6039
> confint(svymean(~PCAR_SUM,design=dclus1sub,na.rm=T))

      2.5 %   97.5 %
PCAR_SUM 0.7081292 0.777222

```

```

> # svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T)      #without
replacement

> # confint(svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))

> svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T)
#without replacement

  GR_ETARIO  PCAR_SUM          se DEff.PCAR_SUM
1          1 0.4738617 0.030676359      2.234487
2          2 0.5986159 0.038137386      3.492764
3          3 0.7189097 0.033171591      3.190877
4          4 0.9702970 0.010085649      1.778827
5          5 0.9912281 0.004576374      1.370524

> confint(svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T))
      2.5 %      97.5 %
1 0.4137372 0.5339863
2 0.5238680 0.6733638
3 0.6538946 0.7839248
4 0.9505295 0.9900645
5  0.9822585 1.0001976

```

2.3 Sem peso, sem estrato e cluster (Amostra Aleatória Simples com reposição)

```

> svymean(~CEOCPOD, design=dclus1sub, deff="replace", na.rm=T)

      mean      SE  DEff
CEOCPOD 10.36357  0.21881 0.9998

> confint(svymean(~CEOCPOD, design=dclus1sub, na.rm=T))

      2.5 %   97.5 %
CEOCPOD 9.934721 10.79242

> svyby(~CEOCPOD, ~GR_ETARIO, design=dclus1sub, svymean, deff="TRUE", na.rm=T) #without
replacement

      GR_ETARIO  CEOCPOD      se DEff.CEOCPOD
1           1  2.028668 0.1260300      Inf
2           2  2.121107 0.1140064      Inf
3           3  4.047700 0.1712273      Inf
4           4 15.853465 0.3256427      Inf
5           5 29.033333 0.2439895      Inf

> confint(svyby(~CEOCPOD, ~GR_ETARIO, design=dclus1sub, svymean, deff="TRUE", na.rm=T))

      2.5 %   97.5 %
1  1.781653  2.275682
2  1.897659  2.344556
3  3.712101  4.383299
4 15.215217 16.491713
5 28.555123 29.511544

> # svyby(~CEOCPOD, ~GR_ETARIO, design=dclus1sub, svymean, deff="replace", na.rm=T)
#without replacement

> # confint(svyby(~CEOCPOD, ~GR_ETARIO, design=dclus1sub, svymean, deff="replace", na.rm=T))
>
> #Prop:
> svymean(~PCAR_SUM, design=dclus1sub, deff="replace", na.rm=T)

      mean      SE  DEff
PCAR_SUM 0.742676 0.008214 0.9998

> confint(svymean(~PCAR_SUM, design=dclus1sub, na.rm=T))

      2.5 %   97.5 %
PCAR_SUM 0.7265765 0.7587747

> # svyby(~PCAR_SUM, ~GR_ETARIO, design=dclus1sub, svymean, deff="TRUE", na.rm=T) #without
replacement

> # confint(svyby(~PCAR_SUM, ~GR_ETARIO, design=dclus1sub, svymean, deff="TRUE", na.rm=T))

> svyby(~PCAR_SUM, ~GR_ETARIO, design=dclus1sub, svymean, deff="replace", na.rm=T)
#without replacement

      GR_ETARIO  PCAR_SUM      se DEff.PCAR_SUM

```

```
1      1 0.4738617 0.020506168    0.9984796
2      2 0.5986159 0.020390430    0.9984359
3      3 0.7189097 0.018555692    0.9984624
4      4 0.9702970 0.007555143    0.9981857
5      5 0.9912281 0.003906007    0.9984116

> confint(svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub, svymean, deff="replace", na.rm=T))
      2.5 %    97.5 %
1 0.4336704 0.5140531
2 0.5586514 0.6385804
3 0.6825412 0.7552782
4 0.9554892 0.9851048
5 0.9835724 0.9988837
```


3 Interior II

3.1 Com plano amostral complexo e com pesos

```

> #Mean:

> #svyby(~CEO,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T)      #without
replacement

> #confint(svyby(~CEO,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))

> #svyby(~CPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T)      #without
replacement

> #confint(svyby(~CPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))

> svymean(~CEOCPOD, design=dclus1sub, deff="replace",na.rm=T)

      mean      SE  DEff
CEOCPOD 12.81713  0.87983 12.666

> confint(svymean(~CEOCPOD,design=dclus1sub,na.rm=T))

      2.5 %    97.5 %
CEOCPOD 11.0927 14.54156

> svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T)      #without
replacement

      GR_ETARIO  CEOCPOD      se DEff.CEOCPOD
1           1  2.226055 0.2321308    2.242154
2           2  2.734540 0.3441996    4.826674
3           3  4.541977 0.4691112    5.441413
4           4 15.292134 1.5332716   12.919842
5           5 28.932977 0.4887189    2.471892

> confint(svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))

      2.5 %    97.5 %
1  1.771087  2.681023
2  2.059921  3.409159
3  3.622535  5.461418
4 12.286977 18.297291
5 27.975105 29.890848

> # svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T)
#without replacement

> # confint(svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T))
>

> #Prop:

> svymean(~PCAR_SUM, design=dclus1sub, deff="replace",na.rm=T)

      mean      SE  DEff
PCAR_SUM 0.861517 0.017229 5.1355

> confint(svymean(~PCAR_SUM,design=dclus1sub,na.rm=T))

      2.5 %    97.5 %

```

```
PCAR_SUM 0.8277488 0.8952859
```

```
> # svyby(~PCAR_SUM,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T) #without replacement
```

```
> # confint(svyby(~PCAR_SUM,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T))
```

```
> svyby(~PCAR_SUM,~GR_ETARIO, design=dcluslsub,svymean, deff="replace",na.rm=T) #without replacement
```

	GR_ETARIO	PCAR_SUM	se	DEff.PCAR_SUM
1	1	0.5235006	0.035608543	2.028158
2	2	0.6688565	0.032285939	1.769561
3	3	0.7770328	0.038110820	3.898246
4	4	0.9497156	0.023910533	5.279472
5	5	0.9951771	0.004613796	1.680935

```
> confint(svyby(~PCAR_SUM,~GR_ETARIO, design=dcluslsub,svymean, deff="replace",na.rm=T))
```

```
2.5 % 97.5 %
```

```
1 0.4537092 0.5932921
```

```
2 0.6055773 0.7321358
```

```
3 0.7023370 0.8517286
```

```
4 0.9028518 0.9965794
```

```
5 0.9861343 1.0042200
```

3.2 Com plano amostral complexo e sem pesos

```

> #Mean:

> #svyby(~CEO,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T)      #without
replacement

> #confint(svyby(~CEO,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))

> #svyby(~CPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T)      #without
replacement

> #confint(svyby(~CPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))

> svymean(~CEOCPOD, design=dclus1sub, deff="replace",na.rm=T)

      mean      SE  DEff
CEOCPOD 10.64891  0.79102  9.799

> confint(svymean(~CEOCPOD,design=dclus1sub,na.rm=T))

      2.5 %    97.5 %
CEOCPOD 9.098546 12.19928

> svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T)      #without
replacement

      GR_ETARIO  CEOCPOD      se DEff.CEOCPOD
1          1    2.317500 0.1793918    1.233888
2          2    2.641910 0.2754185    3.419420
3          3    4.648069 0.4389106    4.696598
4          4   15.606335 1.2488186    8.547056
5          5   28.955263 0.4292671    1.970213

> confint(svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))

      2.5 %    97.5 %
1  1.965899  2.669101
2  2.102099  3.181720
3  3.787820  5.508318
4 13.158695 18.053974
5 28.113915 29.796611

> # svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T)
#without replacement

> # confint(svyby(~CEOCPOD,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T))
>

> #Prop:

> svymean(~PCAR_SUM, design=dclus1sub, deff="replace",na.rm=T)

      mean      SE  DEff
PCAR_SUM 0.790315 0.017039 3.6161

> confint(svymean(~PCAR_SUM,design=dclus1sub,na.rm=T))

      2.5 %    97.5 %
PCAR_SUM 0.7569184 0.8237111

```

```

> # svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T)      #without
replacement

> # confint(svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))

> svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T)
#without replacement

  GR_ETARIO  PCAR_SUM          se DEff.PCAR_SUM
1          1 0.5275000 0.027966897      1.2520947
2          2 0.6684350 0.034332710      1.9997521
3          3 0.7939914 0.030714596      2.6818931
4          4 0.9502262 0.020626273      3.9669132
5          5 0.9973684 0.002582568      0.9630985

> confint(svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T))
      2.5 %      97.5 %
1 0.4726859 0.5823141
2 0.6011441 0.7357259
3 0.7337919 0.8541909
4 0.9097995 0.9906530
5 0.9923067 1.0024302

```

3.3 Sem peso, sem estrato e cluster (Amostra Aleatória Simples com reposição)

```

> #Mean:
> #svyby(~CEO,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T)      #without
replacement
> #confint(svyby(~CEO,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T))
> #svyby(~CPOD,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T)      #without
replacement
> #confint(svyby(~CPOD,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T))
> svymean(~CEOCPOD, design=dcluslsub, deff="replace",na.rm=T)
      mean      SE  DEff
CEOCPOD 10.64891  0.25265 0.9997
> confint(svymean(~CEOCPOD,design=dcluslsub,na.rm=T))
      2.5 %   97.5 %
CEOCPOD 10.15372 11.1441
> svyby(~CEOCPOD,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T)      #without
replacement
      GR_ETARIO  CEOCPOD      se DEff.CEOCPOD
1           1  2.317500 0.1613104      Inf
2           2  2.641910 0.1487584      Inf
3           3  4.648069 0.2023295      Inf
4           4 15.606335 0.4267176      Inf
5           5 28.955263 0.3054500      Inf
> confint(svyby(~CEOCPOD,~GR_ETARIO, design=dcluslsub,svymean, deff="TRUE",na.rm=T))
      2.5 %   97.5 %
1  2.001337  2.633663
2  2.350349  2.933471
3  4.251510  5.044627
4 14.769984 16.442686
5 28.356592 29.553934
> # svyby(~CEOCPOD,~GR_ETARIO, design=dcluslsub,svymean, deff="replace",na.rm=T)
#without replacement
> # confint(svyby(~CEOCPOD,~GR_ETARIO, design=dcluslsub,svymean, deff="replace",na.rm=T))
>
> #Prop:
> svymean(~PCAR_SUM, design=dcluslsub, deff="replace",na.rm=T)
      mean      SE  DEff
PCAR_SUM 0.790315 0.008959 0.9997
> confint(svymean(~PCAR_SUM,design=dcluslsub,na.rm=T))
      2.5 %   97.5 %
PCAR_SUM 0.7727554 0.8078741

```

```

> # svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T)      #without
replacement

> # confint(svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="TRUE",na.rm=T))

> svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T)
#without replacement

  GR_ETARIO  PCAR_SUM          se DEff.PCAR_SUM
1          1 0.5275000 0.024964234      0.9976658
2          2 0.6684350 0.024248191      0.9975133
3          3 0.7939914 0.018736726      0.9980200
4          4 0.9502262 0.010345195      0.9979034
5          5 0.9973684 0.002628333      0.9975342

> confint(svyby(~PCAR_SUM,~GR_ETARIO, design=dclus1sub,svymean, deff="replace",na.rm=T))
      2.5 %      97.5 %
1 0.4785710 0.5764290
2 0.6209094 0.7159606
3 0.7572681 0.8307147
4 0.9299500 0.9705025
5 0.9922170 1.0025199

```